Available online at www.sciencedirect.com

ScienceDirect

Journal of Complexity 23 (2007) 225–244

Journal of COMPLEXITY

www.elsevier.com/locate/jco

Regularization by truncated Cholesky factorization: A comparison of four different approaches☆

Barbara Kaltenbacher

University of Erlangen, Germany

Received 17 March 2006; accepted 27 July 2006
Available online 3 November 2006

Abstract

Due to the principle of regularization by restricting the number of degrees of freedom, truncating the Cholesky factorization of a symmetric positive definite matrix can be expected to have a stabilizing effect. Based on this idea, we consider four different approaches for regularizing ill-posed linear operator equations. Convergence in the noise free case as well as—with an appropriate a priori truncation rule—in the situation of noisy data is analyzed. Moreover, we propose an a posteriori truncation rule and characterize its convergence. Numerical tests illustrate the theoretical results. Both analysis and computations suggest one of the four variants to be favorable to the others.

Keywords: Ill-posed problems; Regularization; Cholesky factorization

1. Introduction

Consider the linear operator equation

$$Tx = y, \tag{1}$$

where $T : l^2 \to l^2$ is a compact linear operator and l^2 is the usual space of quadratically summable sequences with the norm

$$\|v\| = \sqrt{\sum_{j=1}^{\infty} v_j^2}, \quad v = (v_j)_{j\in\mathbb{N}} \in l^2.$$

☆ Supported by the Austrian Academy of Sciences within the Radon Institute for Computational and Applied Mathematics, as well as the German Science Foundation DFG under Grant Ka 1778/1.

E-mail address: barbara@1se.eel.uni-erlangen.de.

0885-064X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jco.2006.07.003

(Note that by means of development with respect to appropriate bases, compact operator equations in arbitrary separable Hilbert spaces can be transferred to the form (1).) In this situation we deal with an ill-posed problem in the sense that x does not depend continuously on the data y, and therefore have to apply some regularization in order to be able to recover a stable approximation to the exact solution also from noisy data y^δ, as it will be given in practice. We here assume that we know the noise level δ in

$$\|y - y^\delta\| \leqslant \delta. \tag{2}$$

Moreover, a solution $x^\dagger$ to (1) is assumed to exist (i.e., $y \in \mathcal{R}(T)$) and to be unique. In order to simplify the exposition, we here even assume that the range $\mathcal{R}(T)$ of T is dense in l^2 and that its nullspace $\mathcal{N}(T)$ is $\{0\}$.

Our aim is to use the principle of regularization by discretization, i.e., by restriction of the degrees of freedom to finitely many (cf., e.g., [7,9], as well as [3, Section 3.3] and the references therein). More precisely, we consider a truncated Cholesky factorization, i.e., one that only takes into account a relatively small number of columns (and rows) in the lower triangular matrix produced by this factorization method. Using a truncated version of the Cholesky factorization can be viewed as a finite rank approximation to the inverse of the forward operator. In this sense, the present paper may be viewed as a pre-study to investigations on regularization by $\mathcal{H}$ matrix approximation (cf., e.g., [1,6].) On the other hand, the results obtained here can be regarded as a first step into the direction of regularization by truncated factorizations of more general nonsymmetric matrices such as LU or QR decompositions.

To be able to make use of Cholesky factorization for equations of the form (1) with not necessarily symmetric positive definite T, we here discuss two main approaches:

The first one starts from the normal equation corresponding to (1)

$$T^{\mathrm{T}} T x = T^{\mathrm{T}} y. \tag{3}$$

Here Cholesky factorization is applied to $T^{\mathrm{T}} T$.

For the second one, we depart from the principle of regularizing by projecting the infinite dimensional equation (1) onto some finite dimensional subspace Y_n of the data space (here l^2)

$$Q_n T x = Q_n y,$$

where Q_n is the orthogonal projection onto Y_n, and use the minimum norm solution of the projected equation, with the given noisy data y^δ inserted in place of y

$$x_n^\delta := (Q_n T)^\dagger Q_n y^\delta \tag{4}$$

as an approximation for $x^\dagger$. Here $B^\dagger$ denotes the generalized inverse of some operator B:

$$B^\dagger : \mathcal{R}(B) \cup \mathcal{R}(B)^\perp \to \mathcal{N}(B)^\perp = \overline{\mathcal{R}(B^T)}, \quad B^\dagger|_{\mathcal{R}(B)} := (B|_{\mathcal{N}(B)^\perp})^{-1},$$
$$B^\dagger|_{\mathcal{R}(B)^\perp} := 0.$$

Therewith x_n^δ has to be contained in $T^{\mathrm{T}} Y_n$, so that it can be written as

$$x_n^\delta = T^{\mathrm{T}} u_n,$$

where

$$Q_n T T^{\mathrm{T}} u_n = Q_n y^\delta. \tag{5}$$

Since TT^{T} is a self adjoint nonnegative definite operator from l^2 to l^2, the idea is now to define the projections Q_n (and therewith the projection spaces Y_n) by applying a truncated Cholesky decomposition to TT^{T}.

In either of the two situations we have the possibility of taking into account

(a) the full first n columns;
(b) only the upper quadratic $n \times n$ part

of the lower triangular matrix produced by the respective Cholesky factorization.

Accordingly, we arrive at altogether four different methods. Considering

$$T^{\mathrm{T}}T = LL^{\mathrm{T}} \tag{6}$$

and the decomposition

$$L = \begin{pmatrix} L_{nn} & 0 \\ L_{rn} & L_{rr} \end{pmatrix} = \begin{pmatrix} L_n & 0 \\ & L_{rr} \end{pmatrix}, \tag{7}$$

with $L_{nn} : \mathbb{R}^n \to \mathbb{R}^n$, $L_{rn} : \mathbb{R}^n \to l^2$, $L_{rr} : l^2 \to l^2$, $L_n = \begin{pmatrix} L_{nn} \\ L_{rn} \end{pmatrix} : \mathbb{R}^n \to l^2$ or

$$TT^{\mathrm{T}} = \tilde{L}\tilde{L}^{\mathrm{T}} \tag{8}$$

and the decomposition

$$\tilde{L} = \begin{pmatrix} \tilde{L}_{nn} & 0 \\ \tilde{L}_{rn} & \tilde{L}_{rr} \end{pmatrix} = \begin{pmatrix} \tilde{L}_n & 0 \\ & \tilde{L}_{rr} \end{pmatrix}, \tag{9}$$

with $\tilde{L}_{nn} : \mathbb{R}^n \to \mathbb{R}^n$, $\tilde{L}_{rn} : \mathbb{R}^n \to l^2$, $\tilde{L}_{rr} : l^2 \to l^2$, $\tilde{L}_n = \begin{pmatrix} \tilde{L}_{nn} \\ \tilde{L}_{rn} \end{pmatrix} : \mathbb{R}^n \to l^2$ and correspondingly for some vector $v \in l^2$

$$v = \begin{pmatrix} v^n \\ v^r \end{pmatrix},$$

with $v^n \in \mathbb{R}^n$, $v^r \in l^2$ we have

Method 1(a):

$$\bar{x}_n^\delta := (L_n L_n^{\mathrm{T}})^\dagger T^{\mathrm{T}} y^\delta,$$

Method 1(b):

$$\bar{z}_n^\delta := \begin{pmatrix} (L_{nn} L_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} T^{\mathrm{T}} y^\delta.$$

Method 2(a):

$$\tilde{x}_n^\delta := T^{\mathrm{T}} (\tilde{L}_n \tilde{L}_n^{\mathrm{T}})^\dagger y^\delta,$$

Method 2(b):

$$\tilde{z}_n^\delta := T^{\mathrm{T}} \begin{pmatrix} (\tilde{L}_{nn} \tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} y^\delta.$$

Note that due to our assumption on the range and nullspace on T, the submatrix L_{nn} ($\tilde{L}_{nn}$) is always regular. It can be computed row wise without having to compute the semi-infinite matrix L_{rn} ($\tilde{L}_{rn}$); increasing n by one, i.e., going from L_{nn} to $L_{n+1,n+1}$, amounts to computing one additional row of $n+1$ entries. Moreover, we have

$$(L_n L_n^{\mathrm{T}})^{\dagger} = L_n (L_n^{\mathrm{T}} L_n)^{-2} L_n^{\mathrm{T}}, \quad (\tilde{L}_n \tilde{L}_n^{\mathrm{T}})^{\dagger} = \tilde{L}_n (\tilde{L}_n^{\mathrm{T}} \tilde{L}_n)^{-2} \tilde{L}_n^{\mathrm{T}}, \tag{10}$$

where $L_n^{\mathrm{T}} L_n$ ($\tilde{L}_n^{\mathrm{T}} \tilde{L}_n$) is invertible due to the fact that for all $v^n \in \mathbb{R}^n$

$$L_n^{\mathrm{T}} L_n v^n = 0 \quad \Leftrightarrow \quad L_n v^n = 0 \quad \Leftrightarrow \quad \Big(L_{nn} v^n = 0 \ \wedge \ L_{rn} v^n = 0 \Big),$$

i.e., $\mathcal{R}(L_n^{\mathrm{T}} L_n)^{\perp} = \mathcal{N}(L_n^{\mathrm{T}} L_n) \subseteq \mathcal{N}(L_{nn}) = \{0\}$. To see that L_{rn}, L_n, $\tilde{L}_{rn}$, $\tilde{L}_n$ in fact map into l^2, consider the identity

$$T^{\mathrm{T}} T = L L^{\mathrm{T}} = L_n L_n^{\mathrm{T}} + \begin{pmatrix} 0 & 0 \\ 0 & L_{rr} L_{rr}^{\mathrm{T}} \end{pmatrix} \tag{11}$$

that by left and right multiplication with an arbitrary vector $v \in L^2$ yields

$$\|Tv\|_{l^2}^2 = \left\| L_n^{\mathrm{T}} v \right\|_{l^2}^2 + \left\| L_{rr}^{\mathrm{T}} v^r \right\|_{l^2}^2 = \left\| L_{nn}^{\mathrm{T}} v^n + L_{rn}^{\mathrm{T}} v^r \right\|_{l^2}^2 + \left\| L_{rr}^{\mathrm{T}} v^r \right\|_{l^2}^2, \tag{12}$$

hence

$$\|L_n\|_{\mathbb{R}^n \to l^2} = \left\| L_n^{\mathrm{T}} \right\|_{l^2 \to \mathbb{R}^n} \leqslant \|T\|_{l^2 \to l^2},$$

$$\|L_{rr}\|_{l^2 \to l^2} = \left\| L_{rr}^{\mathrm{T}} \right\|_{l^2 \to l^2} \leqslant \|T\|_{l^2 \to l^2},$$

$$\|L_{rn}\|_{\mathbb{R}^n \to l^2} = \left\| L_{rn}^{\mathrm{T}} \right\|_{l^2 \to \mathbb{R}^n} \leqslant \|T\|_{l^2 \to l^2},$$

where the last inequality is obtained by taking the supremum over all v with $v^n = 0$ in (12).

It will turn out that although all of these four methods seem to be reasonable at a first glance, only the last one converges unconditionally (cf. Theorem 1). As a matter of fact, Method 2(b) can be written in the form (4) with Q_n being just the projection onto the span of the first n unit vectors. Therefore, for this method, convergence immediately follows from known results on regularization by discretization as outlined, e.g., in [3]. Nevertheless, for the sake of completeness we spend a few lines on the proof of convergence also of this method in Theorem 1.

This paper is organized as follows: Section 2 provides a convergence analysis of Methods 1(a)–2(b) both in the case of exact data and in the situation with noisy right-hand side y^{δ}. For the practically relevant latter setting, an a posteriori truncation rule is investigated in Section 3. The theoretical results are illustrated by numerical tests for four model problems in Section 4. Finally, some conclusions are drawn in Section 5.

2. Convergence

In this section we derive sufficient and necessary conditions for convergence of the four described methods to the exact solution $x^{\dagger}$ of (1), considering first of all noiseless data, i.e., $\delta = 0$,

which leads to respective versions

$$\bar{x}_n := \bar{x}_n^0, \quad \bar{z}_n := \bar{z}_n^0, \quad \tilde{x}_n := \tilde{x}_n^0, \quad \tilde{z}_n := \tilde{z}_n^0,$$

and letting n tend to infinity.

Theorem 1. *For method* 1(a), $\bar{x}_n \to x^\dagger$ *as* $n \to \infty$ *if and only if*

$$\exists C \in \mathbb{R}^+ \quad \forall n \in \mathbb{N} \quad \left\| (L_n L_n^{\mathrm{T}})^\dagger \begin{pmatrix} 0 & 0 \\ 0 & L_{rr} L_{rr}^{\mathrm{T}} \end{pmatrix} \right\| \leqslant C; \tag{13}$$

for method 1(b) $\bar{z}_n \to x^\dagger$ *as* $n \to \infty$ *if and only if*

$$\exists C \in \mathbb{R}^+ \quad \forall n \in \mathbb{N} \quad \left\| L_{rn} L_{nn}^{-1} \right\| \leqslant C; \tag{14}$$

for method 2(a) $\tilde{x}_n \to x^\dagger$ *as* $n \to \infty$ *if and only if*

$$\forall x \in l^2 : \quad \begin{pmatrix} 0 & \tilde{L}_{rr}^{\mathrm{T}} \end{pmatrix} \tilde{L}_n (\tilde{L}_n^{\mathrm{T}} \tilde{L}_n)^{-1} x^n \longrightarrow 0 \quad \text{as } n \to \infty \tag{15}$$

for method 2(b) $\tilde{z}_n \to x^\dagger$ *as* $n \to \infty$ *without any additional conditions on* $\tilde{L}$.

Note that due to the Banach Steinhaus theorem, applied to the operator

$$\begin{pmatrix} 0 & 0 \\ M_n & 0 \end{pmatrix} \quad \text{with } M_n := (0 \ \tilde{L}_{rr}^{\mathrm{T}}) \tilde{L}_n (\tilde{L}_n^{\mathrm{T}} \tilde{L}_n)^{-1}, \tag{16}$$

(15) implies

$$\exists C \in \mathbb{R}^+ \quad \forall n \in \mathbb{N} \quad \left\| \begin{pmatrix} 0 & \tilde{L}_{rr}^{\mathrm{T}} \end{pmatrix} \tilde{L}_n (\tilde{L}_n^{\mathrm{T}} \tilde{L}_n)^{-1} \right\| \leqslant C. \tag{17}$$

Conversely, (17) does *not* imply (15). This can be seen by means of the simple counterexample $M_n : (x_1, \ldots, x_n) \mapsto (0, \ldots, 0, x_1, 0, \ldots)$, shifting x_1 to the $(n+1)$st position and erasing the rest, since for $x^n := (1, 0, \ldots, 0)$, one has $M_n x^n = (0, \ldots, 0, 1, 0, \ldots) \not\to 0$.

Proof. Equivalence of convergence of $\bar{x}_n$ to $x^\dagger$ with (13) can be obtained by

$$\begin{aligned} \left\| \bar{x}_n - x^\dagger \right\| &= \left\| \left(\left(L_n L_n^{\mathrm{T}} \right)^\dagger T^{\mathrm{T}} T - I \right) x^\dagger \right\| \\ &= \sqrt{\left\| Proj_{\mathcal{N}(L_n L_n^{\mathrm{T}})} x^\dagger \right\|^2 + \left\| (L_n L_n^{\mathrm{T}})^\dagger \begin{pmatrix} 0 & 0 \\ 0 & L_{rr} L_{rr}^{\mathrm{T}} \end{pmatrix} x^\dagger \right\|^2}, \end{aligned} \tag{18}$$

where we have used (11) and the fact that $(L_n L_n^{\mathrm{T}})^\dagger L_n L_n^{\mathrm{T}}$ is the projection onto the orthogonal complement of the nullspace of $L_n L_n^{\mathrm{T}}$. The first term on the right-hand side of (18) goes to zero since the spaces $(\mathcal{N}(L_n L_n^{\mathrm{T}})^\perp)_{n \in \mathbb{N}} = (\mathcal{R}(L_n L_n^{\mathrm{T}}))_{n \in \mathbb{N}}$ are nested

$$\begin{aligned} w \in \mathcal{R}(L_n L_n^{\mathrm{T}}) = \mathcal{R}(L_n) &\Rightarrow \left(\exists \, v^n \in \mathbb{R}^n \ : w = L_n v^n \right) \\ &\Rightarrow \left(\exists \, v^{n+1} := \begin{pmatrix} v^n \\ 0 \end{pmatrix} \in \mathbb{R}^{n+1} \ : w = L_{n+1} v^{n+1} \right) \\ &\Rightarrow w \in \mathcal{R}(L_{n+1}) = \mathcal{R}(L_{n+1} L_{n+1}^{\mathrm{T}}) \end{aligned}$$

and their union is dense in $\mathcal{R}(T^{\mathrm{T}}T)$, which we have assumed to be dense in l^2. Therefore, convergence occurs if and only if the second term on the right-hand side of (18) goes to zero, which is equivalent to (13). Namely, if (13) holds, then $\left\|(L_nL_n^{\mathrm{T}})^{\dagger}\begin{pmatrix}0&0\\0&L_{rr}L_{rr}^{\mathrm{T}}\end{pmatrix}x^{\dagger}\right\|\leqslant C\left\|x^{\dagger r}\right\|$, which tends to zero as $n\to\infty$, due to the fact that $x^{\dagger}\in l^2$. On the other hand convergence of the second term on the right-hand side of (18) by the Banach Steinhaus theorem implies (13).

To treat convergence of $\bar{z}_n$ defined in method 1(b) we rewrite

$$\begin{aligned}\left\|\bar{z}_n-x^{\dagger}\right\|&=\left\|\left(\begin{pmatrix}(L_{nn}L_{nn}^{\mathrm{T}})^{-1}&0\\0&0\end{pmatrix}T^{\mathrm{T}}T-I\right)x^{\dagger}\right\|\\&=\left\|\begin{pmatrix}0&L_{nn}^{-\mathrm{T}}L_{rn}^{\mathrm{T}}\\0&-I_r\end{pmatrix}x^{\dagger}\right\|\\&=\sqrt{\left\|L_{nn}^{-\mathrm{T}}L_{rn}^{\mathrm{T}}x^{\dagger\,r}\right\|^2+\left\|x^{\dagger\,r}\right\|^2},\end{aligned}\tag{19}$$

where $A^{-\mathrm{T}}$ abbreviates $(A^{-1})^{\mathrm{T}}=(A^{\mathrm{T}})^{-1}$ and we have inserted

$$T^{\mathrm{T}}T=\begin{pmatrix}L_{nn}L_{nn}^{\mathrm{T}}&L_{nn}L_{rn}^{\mathrm{T}}\\L_{rn}L_{nn}^{\mathrm{T}}&L_{rn}L_{rn}^{\mathrm{T}}+L_{rr}L_{rr}^{\mathrm{T}}\end{pmatrix}.$$

Now, we use the second line in (19) with the Banach Steinhaus theorem for necessity, as well as the fact that $x^{\dagger\,r}\to 0$ as $n\to\infty$ for any $x^{\dagger}\in l^2$ for sufficiency of the uniform boundedness condition (14).

In method 2(a) we have

$$\begin{aligned}\left\|\tilde{x}_n-x^{\dagger}\right\|&=\left\|\left(T^{\mathrm{T}}(\tilde{L}_n\tilde{L}_n^{\mathrm{T}})^{\dagger}T-I\right)x^{\dagger}\right\|\\&=\left\|\left(\tilde{L}^{\mathrm{T}}(\tilde{L}_n\tilde{L}_n^{\mathrm{T}})^{\dagger}\tilde{L}-I\right)\hat{x}\right\|\\&=\left\|\left(\begin{pmatrix}I_n\\M_n\end{pmatrix}\begin{pmatrix}I_n&M_n^{\mathrm{T}}\end{pmatrix}-I\right)\hat{x}\right\|\\&=\sqrt{\left\|M_n^{\mathrm{T}}\hat{x}^r\right\|^2+\left\|M_n\hat{x}^n+(M_nM_n^{\mathrm{T}}-I_r)\hat{x}^r\right\|^2},\end{aligned}\tag{20}$$

with M_n as defined in (16), where we have used (10). Here $\hat{x}$ is chosen such that $\tilde{L}\hat{x}=Tx^{\dagger}$ and $\|\hat{x}\|=\|x^{\dagger}\|$, which is possible due to $\mathcal{R}(T)=\mathcal{R}((TT^{\mathrm{T}})^{\frac{1}{2}})=\mathcal{R}((\tilde{L}\tilde{L}^{\mathrm{T}})^{\frac{1}{2}})=\mathcal{R}(\tilde{L})$ and equality of the singular values of T and $\tilde{L}$ (cf. [3, Proposition 2.18]).

From the third line in (20) we first of all conclude that convergence of $\tilde{x}_n$ for arbitrary $x^{\dagger}$, by the Banach Steinhaus theorem implies uniform boundedness of M_n by some constant C. If this uniform boundedness holds, we can estimate from below according to

$$\left\|\tilde{x}_n-x^{\dagger}\right\|\geqslant\left\|M_n\hat{x}^n\right\|-\max\{C^2,1\}\left\|\hat{x}^r\right\|,$$

which since $\hat{x}^r$ goes to zero as $n\to\infty$ for $\hat{x}\in l^2$ and $x^{\dagger}$ (and therewith $\hat{x}$) was arbitrary (note that due to our assumptions of bijectivity of T and $\tilde{L}$ on their respective ranges, there is a one-to-one correspondence between $x^{\dagger}$ and $\hat{x}$) yields necessity of (15). Sufficiency for convergence immediately follows from the last line in (20) together with (17) and $\hat{x}^r\to 0$ as $n\to\infty$.

Finally, convergence of $\tilde{z}_n$ according to method 2(b) follows from the error representation

$$\begin{aligned}\left\|\tilde{z}_n - x^\dagger\right\| &= \left\|\left(T^{\mathrm{T}}\begin{pmatrix}(\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0\\ 0 & 0\end{pmatrix}T - I\right)x^\dagger\right\| \\ &= \left\|\left(\tilde{L}^{\mathrm{T}}\begin{pmatrix}(\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0\\ 0 & 0\end{pmatrix}\tilde{L} - I\right)\hat{x}\right\| \\ &= \left\|\hat{x}^r\right\|, \end{aligned} \tag{21}$$

with $\hat{x}$ chosen like in the proof for method 2(a) above. □

Now, the influence of noise in the data is taken into account by using an appropriate truncation index $n^* = n^*(\delta)$ in dependence of δ, that guarantees convergence of $\bar{x}_{n^*}^\delta$ ($\bar{z}_{n^*}^\delta$, $\tilde{x}_{n^*}^\delta$, $\tilde{z}_{n^*}^\delta$) to $x^\dagger$ as $\delta \to 0$. It is straightforward to see that the noise amplification factors in the respective methods, i.e. γ_n^{1a}, γ_n^{1b}, γ_n^{2a}, γ_n^{2b} in

$$\left\|\bar{x}_n^\delta - \bar{x}_n\right\| \leqslant \gamma_n^{1a}\delta, \quad \left\|\bar{z}_n^\delta - \bar{z}_n\right\| \leqslant \gamma_n^{1b}\delta,$$
$$\left\|\tilde{x}_n^\delta - \tilde{x}_n\right\| \leqslant \gamma_n^{2a}\delta, \quad \left\|\tilde{z}_n^\delta - \tilde{z}_n\right\| \leqslant \gamma_n^{2b}\delta,$$

with δ the data noise level in (2), are given as follows:
For method 1(a)

$$\gamma_n^{1a} = \left\|(L_nL_n^{\mathrm{T}})^\dagger T^{\mathrm{T}}\right\| = \sqrt{\lambda_{\max}((L_nL_n^{\mathrm{T}})^\dagger LL^{\mathrm{T}}(L_nL_n^{\mathrm{T}})^\dagger)} \leqslant \frac{\sqrt{1+C}}{\sqrt{\lambda_{\min}(L_n^{\mathrm{T}}L_n)}} \tag{22}$$

if (13) holds.
For method 1(b)

$$\begin{aligned}\gamma_n^{1b} &= \left\|\begin{pmatrix}(L_{nn}L_{nn}^{\mathrm{T}})^{-1} & 0\\ 0 & 0\end{pmatrix}T^{\mathrm{T}}\right\| \\ &= \sqrt{\lambda_{\max}\left(\begin{pmatrix}(L_{nn}L_{nn}^{\mathrm{T}})^{-1} & 0\\ 0 & 0\end{pmatrix}LL^{\mathrm{T}}\begin{pmatrix}(L_{nn}L_{nn}^{\mathrm{T}})^{-1} & 0\\ 0 & 0\end{pmatrix}\right)} \\ &= \frac{1}{\sqrt{\lambda_{\min}\left(L_{nn}L_{nn}^{\mathrm{T}}\right)}}.\end{aligned}$$

For method 2(a)

$$\gamma_n^{2a} = \left\|T^{\mathrm{T}}(\tilde{L}_n\tilde{L}_n^{\mathrm{T}})^\dagger\right\| \leqslant \frac{\sqrt{1+C}}{\sqrt{\lambda_{\min}(\tilde{L}_n^{\mathrm{T}}\tilde{L}_n)}}$$

if

$$\exists C \in \mathbb{R}^+ \quad \forall n \in \mathbb{N} \quad \left\|(\tilde{L}_n\tilde{L}_n^{\mathrm{T}})^\dagger\begin{pmatrix}0 & 0\\ 0 & \tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}}\end{pmatrix}\right\| \leqslant C; \tag{23}$$

i.e., a condition similar to (13) but different from the convergence condition (15) for method 2(a) with exact data holds.

For method 2(b)

$$\gamma_n^{2b} = \left\| T^{\mathrm{T}} \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} \right\| = \frac{1}{\sqrt{\lambda_{\min}\left(\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}}\right)}}.$$

Corollary 1. *In the situation of noisy data y^δ satisfying* (2), *let the stopping index $n^* = n^*(\delta)$ in method* 1(a) *be chosen such that*

$$n^*(\delta) \to \infty \quad \text{and} \quad \gamma_{n^*(\delta)}^{1a} \cdot \delta \to 0 \quad \text{as } \delta \to 0, \tag{24}$$

and analogously for methods 1(b), 2(a), 2(b) *with γ_n^{1b}, γ_n^{2a}, γ_n^{2b}. Moreover, let the convergence conditions of Theorem* 1 *and in case of method* 2(a) *additionally* (23) *hold. Then the respective method converges as the noise level tends to zero, i.e.,*

$$\bar{x}_{n^*(\delta)}^\delta \to x^\dagger \quad \text{as } \delta \to 0,$$

and analogously for $\bar{z}_{n^*(\delta)}^\delta$, $\tilde{x}_{n^*(\delta)}^\delta$, $\tilde{z}_{n^*(\delta)}^\delta$.

The truncation index choice (24) is an a priori rule that requires knowledge of positive lower bounds of the eigenvalues appearing in the noise amplification factor estimates. Note that by the Courant–Fisher variational characterization of eigenvalues (cf., e.g. [2, Theorem 8.2]), the following relations between the nth eigenvalue of $T^{\mathrm{T}}T = LL^{\mathrm{T}}$ (ordered in decreasing magnitude) and $\lambda_{\min}(L_{nn}L_{nn}^{\mathrm{T}})$ or $\lambda_{\min}(L_n^{\mathrm{T}}L_n)$, respectively, hold:

$$\begin{aligned}
\lambda_n(LL^{\mathrm{T}}) &= \inf_{\dim(\mathcal{L})=n-1} \sup\{v^{\mathrm{T}}LL^{\mathrm{T}}v \mid \|v\| = 1 \wedge\ v \in \mathcal{L}^\perp\} \\
&\leqslant \inf_{\dim(\mathcal{L})=n-1 \wedge \mathcal{L}^\perp \subseteq \mathrm{span}(e_1,\dots,e_n)} \\
&\quad \times \sup\left\{ \underbrace{v^{\mathrm{T}} \begin{pmatrix} L_{nn}L_{nn}^{\mathrm{T}} & L_{nn}L_{rn}^{\mathrm{T}} \\ L_{rn}L_{nn}^{\mathrm{T}} & L_{rn}L_{rn}^{\mathrm{T}} + L_{rr}L_{rr}^{\mathrm{T}} \end{pmatrix} v}_{= {v^n}^{\mathrm{T}} L_{nn}L_{nn}^{\mathrm{T}} v^n} \;\middle|\; \|v\| = 1,\ v \in \mathcal{L}^\perp \right\} \\
&\leqslant \inf_{\mathcal{L} \subseteq \mathrm{span}(e_1,\dots,e_n) \wedge \dim(\mathcal{L})=n-1 \wedge \mathcal{L}^\perp \subseteq \mathrm{span}(e_1,\dots,e_n)} \\
&\quad \times \sup\{{v^n}^{\mathrm{T}} L_{nn}L_{nn}^{\mathrm{T}} v^n \mid \|v^n\| = 1,\ v^n \in \mathcal{L}^\perp\} \\
&= \lambda_{\min}(L_{nn}L_{nn}^{\mathrm{T}}),
\end{aligned} \tag{25}$$

where $e_1, \dots, e_n$ are the first n unit vectors.

$$\begin{aligned}
\lambda_n(LL^{\mathrm{T}}) &= \inf_{\dim(\mathcal{L})=n-1} \sup\left\{ \underbrace{v^{\mathrm{T}} \left(L_n L_n^{\mathrm{T}} + \begin{pmatrix} 0 & 0 \\ 0 & L_{rr}L_{rr}^{\mathrm{T}} \end{pmatrix} \right) v}_{\geqslant v^{\mathrm{T}} L_n L_n^{\mathrm{T}} v} \;\middle|\; \|v\| = 1 \wedge\ v \in \mathcal{L}^\perp \right\} \\
&\geqslant \lambda_n(L_n L_n^{\mathrm{T}}) = \lambda_{\min}(L_n^{\mathrm{T}} L_n);
\end{aligned} \tag{26}$$

analogously of course for $TT^{\mathrm{T}} = \tilde{L}\tilde{L}^{\mathrm{T}}$.

3. A posteriori truncation: a characterization of convergence with the discrepancy principle

In this section, we derive and analyze an a posteriori truncation index choice, based on the so-called discrepancy principle, that, generally speaking, chooses out of a family of possible regularized approximations, the most stable one such that on the other hand the residual is of the order of magnitude of the data noise. In our context, this reads as

$$n^* = n^*(\delta, y^\delta) = \min\left\{ n \in \mathbb{N} \;\middle|\; \left\| T\bar{x}_n^\delta - y^\delta \right\| \leqslant \tau\delta \right\}, \tag{27}$$

with a fixed constant $\tau > 0$ for method 1(a), and analogously with $\bar{x}_n^\delta$ replaced by $\bar{z}_n^\delta$, $\tilde{x}_n^\delta$, $\tilde{z}_n^\delta$ for methods 1(b), 2(a), and 2(b), respectively. It can be shown, that the conditions

$$\exists C_1 \in \mathbb{R} \quad \forall n \in \mathbb{N}: \quad \lambda_{\max}(L_{rr}L_{rr}^{\mathrm{T}}) \leqslant C_1 \lambda_{\min}(L_{n+1}^{\mathrm{T}} L_{n+1}) \tag{28}$$

for method 1(a),

$$\exists C_1 \in \mathbb{R} \quad \forall n \in \mathbb{N}: \quad \lambda_{\max}(L_{rr}L_{rr}^{\mathrm{T}}) \leqslant C_1 \lambda_{\min}(L_{n+1,n+1} L_{n+1,n+1}^{\mathrm{T}}) \tag{29}$$

for method 1(b) and likewise for $\tilde{L}$, on the eigenvalues of the subblocks of L characterize convergence and, in cases 2(a), 2(b), even optimality of the regularization methods described in the previous section, when combined with a discrepancy principle for the choice of n^*.

First of all we consider the approach via the normal equation:

Theorem 2. *If* (28), *and* (13) *as well as*

$$\exists C^{1a} \in \mathbb{R} \quad \forall n \in \mathbb{N} \quad \left\| (0 \; L_{rr}^{\mathrm{T}}) L_n (L_n^{\mathrm{T}} L_n)^{-1} \right\| \leqslant C^{1a}; \tag{30}$$

hold then method 1(a) *with* (27) *and* $\tau > \max\{C^{1a}, 1\}$, *is a regularization method in the sense that for all* $x^\dagger \in l^2$ *and for all* $y^\delta \in l^2$ *such that* (2) *holds for* $y = Tx^\dagger$,

$$\bar{x}_{n^*}^\delta \to x^\dagger \quad \text{as } \delta \to 0.$$

The same holds true for method 1(b) *if* (29) *and* (14), *as well as* $\tau > 1$ *hold.*

Proof. Consider first of all method 1(a). The residual can be decomposed as

$$T\bar{x}_n^\delta - y^\delta = \underbrace{T(\bar{x}_n^\delta - \bar{x}_n) - (y^\delta - y)}_{(*)} + T(\bar{x}_n - x^\dagger), \tag{31}$$

where the norm of term $(*)$ can be estimated by

$$\left\| T(\bar{x}_n^\delta - \bar{x}_n) - (y^\delta - y) \right\| = \left\| (T(L_n L_n^{\mathrm{T}})^\dagger T^{\mathrm{T}} - I)(y^\delta - y) \right\| \leqslant \max\{C^{1a}, 1\}\delta \tag{32}$$

which can be obtained similarly to (20).
Now we recall (cf. (22)) that under condition (13)

$$\left\| \bar{x}_{n^*}^\delta - x^\dagger \right\| \leqslant \sqrt{1 + C} \frac{\delta}{\sqrt{\lambda_{\min}(L_{n^*}^{\mathrm{T}} L_{n^*})}} + \left\| \bar{x}_{n^*} - x^\dagger \right\| \tag{33}$$

holds, and consider sequences $(\delta_k)_{k\in\mathbb{N}}$, $(y_k)_{k\in\mathbb{N}}$ with $\delta_k \overset{k\to\infty}{\to} 0$, $\|y_k - y\| \leqslant \delta_k$, and $n_k^* := n^*(\delta_k)$ chosen according to the discrepancy principle.

In case $(n_k^*)_{k\in\mathbb{N}}$ has a finite accumulation point, there exists a subsequence $(n_{k_l}^*)_{l\in\mathbb{N}}$ of $(n_k^*)_{k\in\mathbb{N}}$, (which, for simplicity we denote by $(n_l^*)_{l\in\mathbb{N}}$,) that converges to some $N^* \in \mathbb{N}$, so that for all sufficiently large l we have $n_l^* = N^*$ and therewith

$$\left\| \bar{x}_{n_l^*}^{\delta_l} - x^\dagger \right\| \leqslant \sqrt{1+C}\, \underbrace{\frac{\delta_l}{\sqrt{\lambda_{\min}(L_{N^*}^{\mathrm{T}} L_{N^*})}}}_{\to 0 \text{ as } l\to\infty} + \left\| \bar{x}_{N^*} - x^\dagger \right\|. \tag{34}$$

Due to the discrepancy principle and by (31), (32) we have

$$\tau\delta_l \geqslant \left\| T\bar{x}_{n_l^*}^{\delta_l} - y^{\delta_l} \right\| \geqslant \left\| T(\bar{x}_{N^*} - x^\dagger) \right\| - \max\{C^{1a}, 1\}\delta_l.$$

Taking the limit $l \to \infty$ on both sides of this inequality yields

$$T(\bar{x}_{N^*} - x^\dagger) = 0,$$

which due to our assumption that the nullspace of T is trivial, implies

$$\bar{x}_{N^*} = x^\dagger,$$

whence (34) yields convergence of $\bar{x}_{n_l^*}^{\delta_l}$ to $x^\dagger$ as $l \to \infty$.

In the complementary case of $n_k^* \to \infty$ as $k \to \infty$ we can use minimality of n^* in the discrepancy principle and (31), (32) to conclude for $n < n_k^*$.

$$\begin{aligned}
&(\tau - \max\{C^{1a}, 1\})\delta_k \\
&\quad\leqslant \left\| T(\bar{x}_n - x^\dagger) \right\| \\
&\quad= \left\| T\Big((L_n L_n^{\mathrm{T}})^\dagger T^{\mathrm{T}} T - I\Big) x^\dagger \right\| \\
&\quad= \left\| L^{\mathrm{T}}\Big((L_n L_n^{\mathrm{T}})^\dagger L L^{\mathrm{T}} - I\Big) x^\dagger \right\| \\
&\quad= \left\| L^{\mathrm{T}}\Big(-Proj_{\mathcal{N}(L_n L_n^{\mathrm{T}})} + (L_n L_n^{\mathrm{T}})^\dagger \begin{pmatrix} 0 & 0 \\ 0 & L_{rr} L_{rr}^{\mathrm{T}} \end{pmatrix}\Big) x^\dagger \right\| \\
&\quad\leqslant \sqrt{\left\| L L^{\mathrm{T}} Proj_{\mathcal{N}(L_n L_n^{\mathrm{T}})} \right\|}\, \left\| Proj_{\mathcal{N}(L_n L_n^{\mathrm{T}})} x^\dagger \right\| \\
&\qquad + \sqrt{\left\| L L^{\mathrm{T}} (L_n L_n^{\mathrm{T}})^\dagger \begin{pmatrix} 0 & 0 \\ 0 & L_{rr} L_{rr}^{\mathrm{T}} \end{pmatrix} \right\|} \sqrt{\left\| (L_n L_n^{\mathrm{T}})^\dagger \begin{pmatrix} 0 & 0 \\ 0 & L_{rr} L_{rr}^{\mathrm{T}} \end{pmatrix} \right\|}\, \left\| x^{\dagger r} \right\|.
\end{aligned} \tag{35}$$

Here we insert once more $L L^{\mathrm{T}} = L_n L_n^{\mathrm{T}} + \begin{pmatrix} 0 & 0 \\ 0 & L_{rr} L_{rr}^{\mathrm{T}} \end{pmatrix}$ to obtain

$$\left\| L L^{\mathrm{T}} Proj_{\mathcal{N}(L_n L_n^{\mathrm{T}})} \right\| = \left\| \begin{pmatrix} 0 & 0 \\ 0 & L_{rr} L_{rr}^{\mathrm{T}} \end{pmatrix} Proj_{\mathcal{N}(L_n L_n^{\mathrm{T}})} \right\| \leqslant \lambda_{\max}(L_{rr} L_{rr}^{\mathrm{T}})$$

and, by (13),

$$\begin{aligned}&\left\| LL^{\mathrm{T}}(L_nL_n^{\mathrm{T}})^{\dagger}\begin{pmatrix}0&0\\0&L_{rr}L_{rr}^{\mathrm{T}}\end{pmatrix}\right\|\\&\quad=\left\| Proj_{\mathcal{N}(L_nL_n^{\mathrm{T}})^{\perp}}\begin{pmatrix}0&0\\0&L_{rr}L_{rr}^{\mathrm{T}}\end{pmatrix}+\begin{pmatrix}0&0\\0&L_{rr}L_{rr}^{\mathrm{T}}\end{pmatrix}(L_nL_n^{\mathrm{T}})^{\dagger}\begin{pmatrix}0&0\\0&L_{rr}L_{rr}^{\mathrm{T}}\end{pmatrix}\right\|\\&\quad\leqslant(1+C)\lambda_{\max}(L_{rr}L_{rr}^{\mathrm{T}})\end{aligned}$$

so that (35) implies

$$\delta_k\leqslant\frac{1}{\tau-\max\{C^{1a},1\}}\sqrt{\lambda_{\max}(L_{rr}L_{rr}^{\mathrm{T}})}\left(\left\|Proj_{\mathcal{N}(L_nL_n^{\mathrm{T}})}x^{\dagger}\right\|+\sqrt{1+C}\left\|x^{\dagger\,r}\right\|\right).\tag{36}$$

Inserting this, with $n:=n_k^*-1$, in its turn, into (33), and using (28), we obtain

$$\begin{aligned}\left\|\bar{x}_{n_k^*}^{\delta}-x^{\dagger}\right\|&\leqslant\frac{\sqrt{(1+C)C_1}}{\tau-\max\{C^{1a},1\}}\left(\left\|Proj_{\mathcal{N}(L_{n_k^*-1}L_{n_k^*-1}^{\mathrm{T}})}x^{\dagger}\right\|+\sqrt{1+C}\left\|x^{\dagger\,r_k^*+1}\right\|\right)\\&\quad+\left\|\bar{x}_{n_k^*}-x^{\dagger}\right\|\\&\overset{k\to\infty}{\longrightarrow}0\end{aligned}$$

since n_k^* tends to infinity. Now a subsequence–subsequence argument yields the assertion.

The proof for method 1(b) goes analogously with (32) replaced by

$$\left\|T(\bar{z}_n^{\delta}-\bar{z}_n)-(y^{\delta}-y)\right\|=\left\|\left(T\begin{pmatrix}(L_{nn}L_{nn}^{\mathrm{T}})^{-1}&0\\0&0\end{pmatrix}T^{\mathrm{T}}-I\right)(y^{\delta}-y)\right\|\leqslant\delta$$

(which can be shown like (21)), with (33) replaced by

$$\left\|\bar{z}_{n^*}^{\delta}-x^{\dagger}\right\|\leqslant\frac{\delta}{\sqrt{\lambda_{\min}(L_{n^*n^*}L_{n^*n^*}^{T})}}+\left\|\bar{z}_{n^*}-x^{\dagger}\right\|,$$

and with (35), (36) replaced by

$$\begin{aligned}(\tau-1)\delta_k&\leqslant\left\|T(\bar{z}_n-x^{\dagger})\right\|\\&=\left\|T\left(\begin{pmatrix}(L_{nn}L_{nn}^{\mathrm{T}})^{-1}&0\\0&0\end{pmatrix}T^{\mathrm{T}}T-I\right)x^{\dagger}\right\|\\&=\left\|L^{\mathrm{T}}\left(\begin{pmatrix}(L_{nn}L_{nn}^{\mathrm{T}})^{-1}&0\\0&0\end{pmatrix}LL^{\mathrm{T}}-I\right)x^{\dagger}\right\|\\&=\left\|\begin{pmatrix}0&0\\0&L_{rr}^{\mathrm{T}}\end{pmatrix}x^{\dagger}\right\|.\qquad\square\end{aligned}$$

In the approach via projection in data space, where (28), (29) become

$$\exists C_2\in\mathbb{R}\quad\forall n\in\mathbb{N}:\quad\lambda_{\max}(\tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}})\leqslant C_2\lambda_{\min}(\tilde{L}_{n+1}^{\mathrm{T}}\tilde{L}_{n+1})\tag{37}$$

$$\exists C_2\in\mathbb{R}\quad\forall n\in\mathbb{N}:\quad\lambda_{\max}(\tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}})\leqslant C_2\lambda_{\min}(\tilde{L}_{n+1,n+1}\tilde{L}_{n+1,n+1}^{\mathrm{T}})\tag{38}$$

for $TT^{\mathrm{T}} = \tilde{L}\tilde{L}^{\mathrm{T}}$, one can make use of the fact that $\tilde{x}_n^\delta$ and $\tilde{z}_n^\delta$ are in the range of T^{T} to even prove optimal convergence rates i.e., a source condition

$$x^\dagger = (T^{\mathrm{T}}T)^\nu w \tag{39}$$

for some $w \in l^2$, with $0 < \nu \leqslant \frac{1}{2}$ implies the convergence rate $O(\delta^{\frac{2\nu}{2\nu+1}})$.

Theorem 3. *If* (37), *and* (23) *hold then method* 2(a), *with the discrepancy principle and* $\tau > \max\{C, 1\}$, *converges, i.e., for all* $x^\dagger \in l^2$ *and for all* $y^\delta \in l^2$ *such that* (2) *holds for* $y = Tx^\dagger$,

$$\tilde{x}_{n^*}^\delta \to x^\dagger \quad as\ \delta \to 0.$$

The same holds true for method 2(b) *if* (38) *and*

$$\exists C^{2b} \in \mathbb{R} \quad \forall n \in \mathbb{N} \quad \left\| \tilde{L}_{rn}\tilde{L}_{nn}^{-1} \right\| \leqslant C^{2b}, \tag{40}$$

as well as $\tau > \sqrt{1 + (C^{2b})^2}$.
In both cases convergence is order optimal, i.e., for all $\nu \leqslant \frac{1}{2}$ *a source condition* (39) *implies*

$$\left\| \tilde{x}^\delta_{n^*(\delta,y^\delta)} - x^\dagger \right\| \leqslant C\, \|w\|^{\frac{1}{2\nu+1}}\, \delta^{\frac{2\nu}{2\nu+1}} \quad and \quad \left\| \tilde{z}^\delta_{n^*(\delta,y^\delta)} - x^\dagger \right\| \leqslant C\, \|w\|^{\frac{1}{2\nu+1}}\, \delta^{\frac{2\nu}{2\nu+1}}, \tag{41}$$

respectively.

Note that the convergence conditions as well as results for Method 2(b) can be directly deduced from [4, Theorem 1], when viewing this method as a special case of (4) with $Q_n = Proj_{\mathrm{span}(e_1,\dots,e_n)}$.

Proof. We here mainly consider Method 2(b) since in view of Theorem 1 it seems to be the best one at least from the point of view of convergence with exact data. The proof goes analogously for method 2(a); points where differences in the proof would occur are indicated by remarks in brackets.

To show sufficiency of (38) for convergence, we first of all consider the case $\nu = \frac{1}{2}$ in the source condition (39), which is equivalent to

$$x^\dagger = T^{\mathrm{T}} w \tag{42}$$

for some $w \in l^2$ and prove that then

$$\left\| \tilde{z}^\delta_{n^*(\delta,y^\delta)} - x^\dagger \right\| \leqslant \tilde{C}\sqrt{\|w\|}\sqrt{\delta} \tag{43}$$

holds for some constant $\tilde{C} > 0$. Using an argument by Plato (cf. [8]), we can then conclude convergence for any $x^\dagger \in l^2$ (without source condition) and optimality for all $\nu \leqslant \frac{1}{2}$. From the definition of $\tilde{z}_n^\delta$ and the source condition (42), we get

$$\tilde{z}_n^\delta - x^\dagger = T^{\mathrm{T}}(u_n - w)$$

with $u_n = \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} y^\delta$ (in method 2(a), $u_n = (\tilde{L}_n\tilde{L}_n^{\mathrm{T}})^\dagger y^\delta$), so that by the interpolation inequality we can estimate the error for $n = n^*$ as follows

$$\left\| \tilde{z}^\delta_{n^*} - x^\dagger \right\| \leqslant \sqrt{\left\| T(\tilde{z}^\delta_{n^*} - x^\dagger) \right\|}\sqrt{\|u_{n^*} - w\|}. \tag{44}$$

The first term on the right-hand side of (44) can be estimated by

$$\left\| T(\tilde{z}^{\delta}_{n^*} - x^{\dagger}) \right\| = \left\| T\tilde{z}^{\delta}_{n^*} - y \right\| \leqslant \left\| T\tilde{z}^{\delta}_{n^*} - y^{\delta} \right\| + \delta \leqslant (\tau + 1)\delta, \tag{45}$$

where we have used the discrepancy principle. To estimate the second term, we first of all rewrite it as

$$\begin{aligned} \|u_{n^*} - w\| &= \left\| \begin{pmatrix} (\tilde{L}_{n^*n^*}\tilde{L}^{\mathrm{T}}_{n^*n^*})^{-1} & 0 \\ 0 & 0 \end{pmatrix} y^{\delta} - w \right\| \\ &= \left\| \left(\begin{pmatrix} (\tilde{L}_{n^*n^*}\tilde{L}^{\mathrm{T}}_{n^*n^*})^{-1} & 0 \\ 0 & 0 \end{pmatrix} TT^{\mathrm{T}} - I \right) w \right. \\ &\quad \left. + \begin{pmatrix} (\tilde{L}_{n^*n^*}\tilde{L}^{\mathrm{T}}_{n^*n^*})^{-1} & 0 \\ 0 & 0 \end{pmatrix} (y^{\delta} - y) \right\| \\ &\leqslant \sqrt{1 + (C^{2b})^2}\, \|w\| + \frac{\delta}{\lambda_{\min}(\tilde{L}_{n^*n^*}\tilde{L}^{\mathrm{T}}_{n^*n^*})}, \end{aligned} \tag{46}$$

since

$$\begin{aligned} \left\| \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}^{\mathrm{T}}_{nn})^{-1} & 0 \\ 0 & 0 \end{pmatrix} TT^{\mathrm{T}} - I \right\| &= \left\| \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}^{\mathrm{T}}_{nn})^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tilde{L}\tilde{L}^{\mathrm{T}} - I \right\| \\ &= \left\| \begin{pmatrix} 0 & \tilde{L}^{-\mathrm{T}}_{nn}\tilde{L}^{\mathrm{T}}_{rn} \\ 0 & -I_r \end{pmatrix} \right\| \leqslant \sqrt{1 + (C^{2b})^2}. \end{aligned} \tag{47}$$

(For method 2(a)) we get, in place of (46), $\|u_{n^*} - w\| \leqslant \sqrt{1 + C^2}\, \|w\| + \frac{\delta}{\lambda_{\min}(\tilde{L}^{\mathrm{T}}_{n^*}\tilde{L}_{n^*})}$, since

$\left\| (\tilde{L}_n\tilde{L}^{\mathrm{T}}_n)^{\dagger} TT^{\mathrm{T}} - I \right\| \leqslant \sqrt{1 + C^2}$, cf. (18).)

To obtain (43) from (44), (46), we now have to be able to estimate $\frac{\delta}{\lambda_{\min}(\tilde{L}_{n^*n^*}\tilde{L}^{\mathrm{T}}_{n^*n^*})}$ from above by a multiple of $\|w\|$. (Note that then also the respective estimate for method 2(a) holds automatically, by (25), (26).) For this purpose, we use the maximality of n^* in (27) to conclude that for all $n < n^*$

$$\begin{aligned} \tau\delta &< \left\| T\tilde{z}^{\delta}_n - y^{\delta} \right\| \\ &= \left\| \left(TT^{\mathrm{T}} \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}^{T}_{nn})^{-1} & 0 \\ 0 & 0 \end{pmatrix} TT^{\mathrm{T}} - TT^{\mathrm{T}} \right) w \right. \\ &\quad \left. + \left(TT^{\mathrm{T}} \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}^{\mathrm{T}}_{nn})^{-1} & 0 \\ 0 & 0 \end{pmatrix} - I \right) (y^{\delta} - y) \right\| \\ &\leqslant \left\| \begin{pmatrix} 0 & 0 \\ 0 & \tilde{L}_{rr}\tilde{L}^{\mathrm{T}}_{rr} \end{pmatrix} w \right\| + \sqrt{1 + (C^{2b})^2}\delta, \end{aligned} \tag{48}$$

where we have used (47). Inequality (48) implies

$$(\tau - \sqrt{1+(C^{2b})^2})\delta < \left\| \begin{pmatrix} 0 & 0 \\ 0 & \tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}} \end{pmatrix} w \right\| \leqslant \lambda_{\max}(\tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}})\,\|w\|\,,$$

and now the eigenvalue condition (38) with $n := n^* - 1$ comes into play and yields

$$\frac{\delta}{\lambda_{\min}(\tilde{L}_{n^*n^*}\tilde{L}_{n^*n^*}^{\mathrm{T}})} \leqslant \frac{C_2}{\tau - \sqrt{1+(C^{2b})^2}}\,\|w\|\,,$$

which, when inserted into (44),(46) gives (43).

The above mentioned result from [8] yields the conclusions on (optimal) convergence. □

From the proofs of Theorems 2 and 3 it is obvious that the eigenvalue conditions (28), (29), (37), and (38) should be even necessary for convergence of the respective methods with the discrepancy principle. We refer to cf. [5] for a more detailed justification of this conjecture in case of Method 2(b).

Some comments on the eigenvalue conditions (28), (29), (37), (38) are in order:

In the context of method 2(b), note that similarly to (25) one gets

$$\begin{aligned} \lambda_{n+1}(TT^{\mathrm{T}}) &= \inf_{\dim(\mathcal{L})=n} \sup\{v^{\mathrm{T}}TT^{\mathrm{T}}v \mid \|v\| = 1 \wedge\ v \in \mathcal{L}^{\perp}\} \\ &\leqslant \sup\{v^{\mathrm{T}}\tilde{L}\tilde{L}^{\mathrm{T}}v \mid \|v\| = 1 \wedge \forall w^n \in \mathbb{R}^n : v^{\mathrm{T}}\tilde{L}_n w^n = 0\} \\ &\leqslant \sup\{v^{r\,\mathrm{T}}\tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}}v^r \mid \|v\| = 1\} \leqslant \lambda_{\max}(\tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}}), \end{aligned} \tag{49}$$

where we have set $\mathcal{L}$ equal to the span of the n columns of $\tilde{L}_n$ to obtain the first inequality. In view of (38), this is an estimate "in the wrong direction", though, so that we here still have a gap in the theory even for method 2(b) that needed no assumptions on T for convergence in the case of exact data.

With an analogous partition of TT^{T} to the one of $\tilde{L}$

$$TT^{\mathrm{T}} = \begin{pmatrix} A_{nn} & A_{rn}^{\mathrm{T}} \\ A_{rn} & A_{rr} \end{pmatrix} = \begin{pmatrix} \tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}} & \tilde{L}_{nn}\tilde{L}_{rn}^{\mathrm{T}} \\ \tilde{L}_{rn}\tilde{L}_{nn}^{\mathrm{T}} & \tilde{L}_{rn}\tilde{L}_{rn}^{\mathrm{T}} + \tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}} \end{pmatrix}$$

the conditions

$$\forall n \in \mathbb{N}: \quad \lambda_{\max}(A_{rr}) \leqslant C\lambda_{\min}(A_{nn}) \tag{50}$$

and

$$\forall n \in \mathbb{N}: \quad \lambda_{\min}(A_{nn}) \leqslant C_1\lambda_{\min}(A_{n+1,n+1}) \tag{51}$$

are sufficient for (38), (40). We expect that (50), (51) can be achieved by appropriate symmetry preserving column and row reordering strategies.

Note also, that in methods 1(a) and 2(a), the eigenvalue relation (28) or (37) with $n+1$ replaced by n, implies the convergence condition (13), or (15), respectively.

For estimating the truncation index n^* from above, the following corollary applies.

Corollary 2. *Let either*

(i) n^* *be chosen according to the a priori rule* (24) *and the conditions of Corollary* 1 *be satisfied or*

(ii) *n^* be chosen according to the discrepancy principle* (27) *and the conditions of Theorems* 2 *or* 3, *respectively, be satisfied.*

Then

$$\lambda_{\min}(L_{nn}L_{nn}^{\mathrm{T}}) \geqslant \lambda_{n^*}(T^{\mathrm{T}}T) \geqslant \lambda_{\min}(L_nL_n^{\mathrm{T}}) \geqslant C\delta^2 \quad \textit{for method } 1(a)$$

$$\lambda_{\min}(L_{nn}L_{nn}^{\mathrm{T}}) \geqslant C\delta^2 \quad \textit{for method } 1(b)$$

$$\lambda_{\min}(\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}}) \geqslant \lambda_{n^*}(TT^{\mathrm{T}}) \geqslant \lambda_{\min}(\tilde{L}_n\tilde{L}_n^{\mathrm{T}}) \geqslant C\delta^2 \quad \textit{for method } 2(a)$$

$$\lambda_{\min}(\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}}) \geqslant C\delta^2 \quad \textit{for method } 2(b)$$

Proof. In case of the a priori choice, the estimates follow directly from the stopping rule (24) as well as (25), (26).

If the discrepancy principle is used, then the assertions follow from the eigenvalue relations (28), (29), (37), (38) as well as the upper estimates of δ in the proofs of Theorems 2, 3, e.g., (36). Note that in the context of Theorem 3 one can, in place of (48) show such an estimate also in the absence of a source condition via

$$\begin{aligned}
\tau\delta &< \left\| T\tilde{z}_n^{\delta} - y^{\delta} \right\| \\
&= \left\| \left(\begin{pmatrix} TT^{\mathrm{T}}(\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} - I \right) Tx^{\dagger} \right. \\
&\quad \left. + \left(TT^{\mathrm{T}} \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} - I \right) (y^{\delta} - y) \right\| \\
&\leqslant \left\| \left(\tilde{L}\tilde{L}^{\mathrm{T}} \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} - I \right) \tilde{L}\hat{x} \right\| + \sqrt{1 + (C^{2b})^2}\delta \\
&= \left\| \begin{pmatrix} 0 & 0 \\ 0 & \tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}} \end{pmatrix} \hat{x} \right\| + \sqrt{1 + (C^{2b})^2}\delta,
\end{aligned}$$

where $\hat{x} = \tilde{L}^{-1}Tx^{\dagger}$. □

The estimate above implies that for mildly inverse problems, where $\lambda_n(T^{\mathrm{T}}T) \sim n^{\alpha}$ with some $\alpha > 0$, we get an estimate of the form

$$n^* \leqslant C\delta^{-2/\alpha},$$

while for severely ill-posed problems, where $\lambda_n(T^{\mathrm{T}}T) \sim \exp(-\alpha n)$, we obtain

$$n^* \leqslant C + \frac{2}{\alpha}|\log\ \delta|,$$

at least in methods 1(a), 2(a).

To give an overview on the convergence results derived here together with the different necessary and sufficient conditions, we summarize them in the following tables:

Methods		
1(a)	$\bar{x}_n^\delta := (L_n L_n^{\mathrm{T}})^\dagger T^{\mathrm{T}} y^\delta$	$T^{\mathrm{T}}T = LL^{\mathrm{T}}$
1(b)	$\bar{z}_n^\delta := \begin{pmatrix} (L_{nn}L_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} T^{\mathrm{T}} y^\delta$	$T^{\mathrm{T}}T = LL^{\mathrm{T}}$
2(a)	$\bar{x}_n^\delta := T^{\mathrm{T}}(\tilde{L}_n \tilde{L}_n^{\mathrm{T}})^\dagger y^\delta$	$TT^{\mathrm{T}} = \tilde{L}\tilde{L}^{\mathrm{T}}$
2(b)	$\bar{z}_n^\delta := T^{\mathrm{T}} \begin{pmatrix} (\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1} & 0 \\ 0 & 0 \end{pmatrix} y^\delta$	$TT^{\mathrm{T}} = \tilde{L}\tilde{L}^{\mathrm{T}}$

Method	Necessary and sufficient condition
Convergence with exact data ($\delta = 0$)	
1(a)	$\left\| (L_n L_n^{\mathrm{T}})^\dagger \begin{pmatrix} 0 & 0 \\ 0 & L_{rr}L_{rr}^{\mathrm{T}} \end{pmatrix} \right\| \leqslant C$
1(b)	$\left\| L_{rn} L_{nn}^{-1} \right\| \leqslant C$
2(a)	$\left(\begin{pmatrix} 0 & \tilde{L}_{rr}^{\mathrm{T}} \end{pmatrix} \tilde{L}_n (\tilde{L}_n^{\mathrm{T}} \tilde{L}_n)^{-1} \quad 0 \right) \xrightarrow{ptw} 0 \quad \text{as } n \to \infty$
2(b)	–

Method	Truncation rule	Condition additional to case $\delta = 0$
Convergence with $\delta > 0$ and a priori truncation choice		
1(a)	$n^*(\delta) \to \infty$ and $\dfrac{\delta}{\sqrt{\lambda_{\min}(L_{n^*}^{\mathrm{T}} L_{n^*})}} \to 0$ as $\delta \to 0$	–
1(b)	$n^*(\delta) \to \infty$ and $\dfrac{\delta}{\sqrt{\lambda_{\min}\left(L_{n^*n^*} L_{n^*n^*}^{\mathrm{T}}\right)}} \to 0$ as $\delta \to 0$	–
2(a)	$n^*(\delta) \to \infty$ and $\dfrac{\delta}{\sqrt{\lambda_{\min}(\tilde{L}_{n^*}^{\mathrm{T}} \tilde{L}_{n^*})}} \to 0$ as $\delta \to 0$	$\left\| (\tilde{L}_n \tilde{L}_n^{\mathrm{T}})^\dagger \begin{pmatrix} 0 & 0 \\ 0 & \tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}} \end{pmatrix} \right\| \leqslant C$
2(b)	$n^*(\delta) \to \infty$ and $\dfrac{\delta}{\sqrt{\lambda_{\min}\left(\tilde{L}_{n^*n^*} \tilde{L}_{n^*n^*}^{\mathrm{T}}\right)}} \to 0$ as $\delta \to 0$	–

Method	Eigenvalue condition	Condition additional to case $\delta = 0$ and to eigenvalue relation
Convergence with $\delta > 0$ and discrepancy principle		
1(a)	$\dfrac{\lambda_{\max}(L_{rr}L_{rr}^{\mathrm{T}})}{\lambda_{\min}(L_{n+1}^{\mathrm{T}} L_{n+1})} \leqslant C$	$\left\| (0 \; L_{rr}^{\mathrm{T}}) L_n (L_n^{\mathrm{T}} L_n)^{-1} \right\| \leqslant C$
1(b)	$\dfrac{\lambda_{\max}(L_{rr}L_{rr}^{\mathrm{T}})}{\lambda_{\min}(L_{n+1,n+1} L_{n+1,n+1}^{\mathrm{T}})} \leqslant C$	–
2(a)	$\dfrac{\lambda_{\max}(\tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}})}{\lambda_{\min}(\tilde{L}_{n+1}^{\mathrm{T}} \tilde{L}_{n+1})} \leqslant C$	$\left\| (\tilde{L}_n \tilde{L}_n^{\mathrm{T}})^\dagger \begin{pmatrix} 0 & 0 \\ 0 & \tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}} \end{pmatrix} \right\| \leqslant C$
2(b)	$\dfrac{\lambda_{\max}(\tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}})}{\lambda_{\min}(\tilde{L}_{n+1,n+1} \tilde{L}_{n+1,n+1}^{\mathrm{T}})} \leqslant C$	$\left\| \tilde{L}_{rn} \tilde{L}_{nn}^{-1} \right\| \leqslant C$

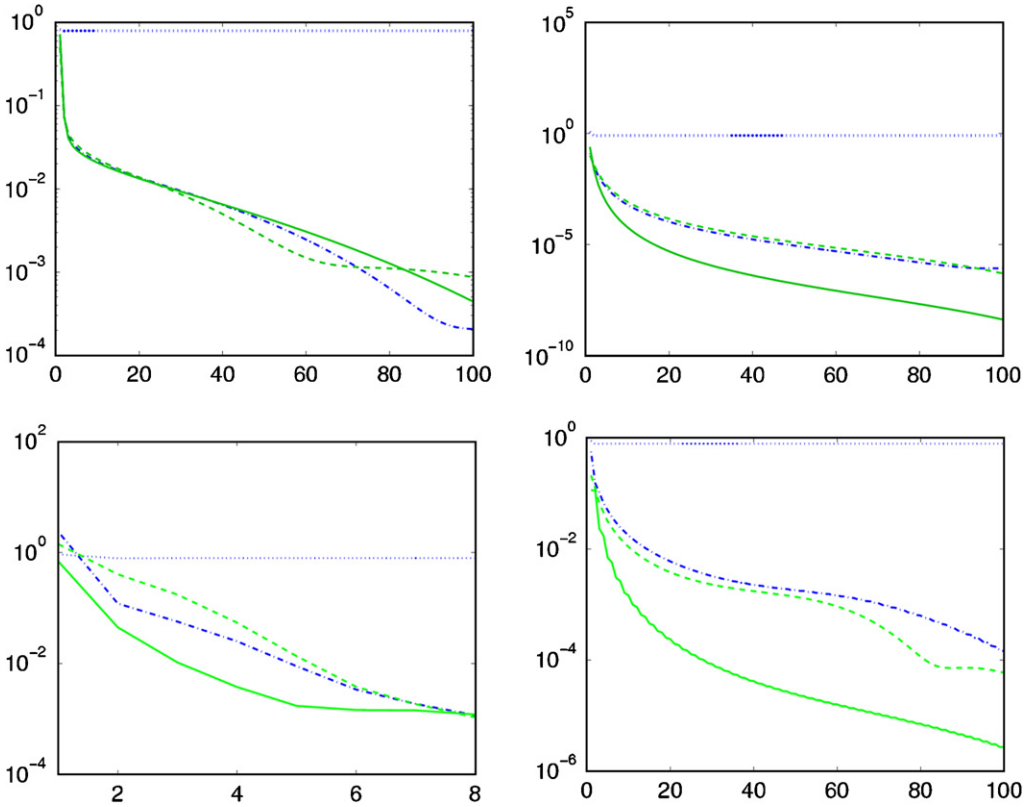

Fig. 1. Error versus truncation index for method 1(a) (dash-dotted), 1(b) (dotted), 2(a) (dashed), and 2(b) (solid) for Example 1 (top left), Example 2 (top right), Example 3 (bottom left), Example 4 (bottom right).

4. Numerical tests

To test the proposed methods, we use the following examples.

Example 1. The Abel integral equation

$$\int_{-1}^{t} \frac{x(s)}{\sqrt{t-s}}\, ds = y(t), \quad t \in (-1, 1), \tag{52}$$

represents the rotational symmetric two-dimensional case in X-ray tomography.

Example 2. Numerical differentiation is a simple but instructive example of a linear ill-posed problem. We here consider twice numerical differentiation $x = y''$ with symmetry boundary conditions $y(-1) = y(1)$ which leads to the integral equation

$$\int_{-1}^{t} (t-s)x(s)\, ds - \frac{1}{2}\int_{-1}^{1} (1-s)x(s)\, ds = y(t), \quad t \in (-1, 1). \tag{53}$$

Table 1
Convergence as $\delta \to 0$ for Example 1 (first block), Example 2 (second block), Example 3 (third block), Example 4 (fourth block)

δ	n^*	$\frac{\|z^\delta_{n^*} - x^\dagger\|}{\|x^\dagger\|}$
1%	17	0.1204
0.5%	41	0.1132
0.25%	58	0.0874
0.125%	84	0.0546
0.8%	3	0.0105
0.4%	4	0.0039
0.2%	4	0.0038
0.1%	5	0.0018
0.1%	8	0.1270
0.05%	11	0.0875
0.025%	12	0.0717
0.0125%	15	0.0483
4%	3	0.0263
2%	3	0.0234
1%	4	0.0178
0.5%	5	0.0070

Example 3. Deconvolution with a Gaussian kernel

$$\int_{-1}^{1} \exp(-100(t-s)^2)x(s)\,ds = y(t), \quad t \in (-1,1) \tag{54}$$

on a compact interval is a severely ill-posed example, due to the smoothness of the integral kernel.

Example 4. Choosing a less smooth kernel, we obtain a convolution integral equation

$$\int_{-1}^{1} sign\left(|t-s| - \frac{1}{2}\right) x(s)\,ds = y(t), \quad t \in (-1,1) \tag{55}$$

that can be expected to be as ill-posed as one differentiation.

In all examples, a development both in preimage and in image space with respect to the orthonormal basis functions $s \mapsto \cos(2\pi ns)$ of $L^2(-1,1)$ leads to a formulation of the problem as an operator equation (1) in l^2.

As exact solution, we used

$$x^\dagger(s) = s^2(1-s)^2$$

and computed the l^2 versions of the respective operators by means of fast Fourier transform. The data were generated synthetically, using a number of nodes (307) in the FFT that is different from the one used for the inverse computations (256) in order to avoid an inverse crime.

To compare the behavior in the noise free case, we display the convergence history with increasing n of each of the four methods in a semi-logarithmic plot in Fig. 1. Here, the unconditional

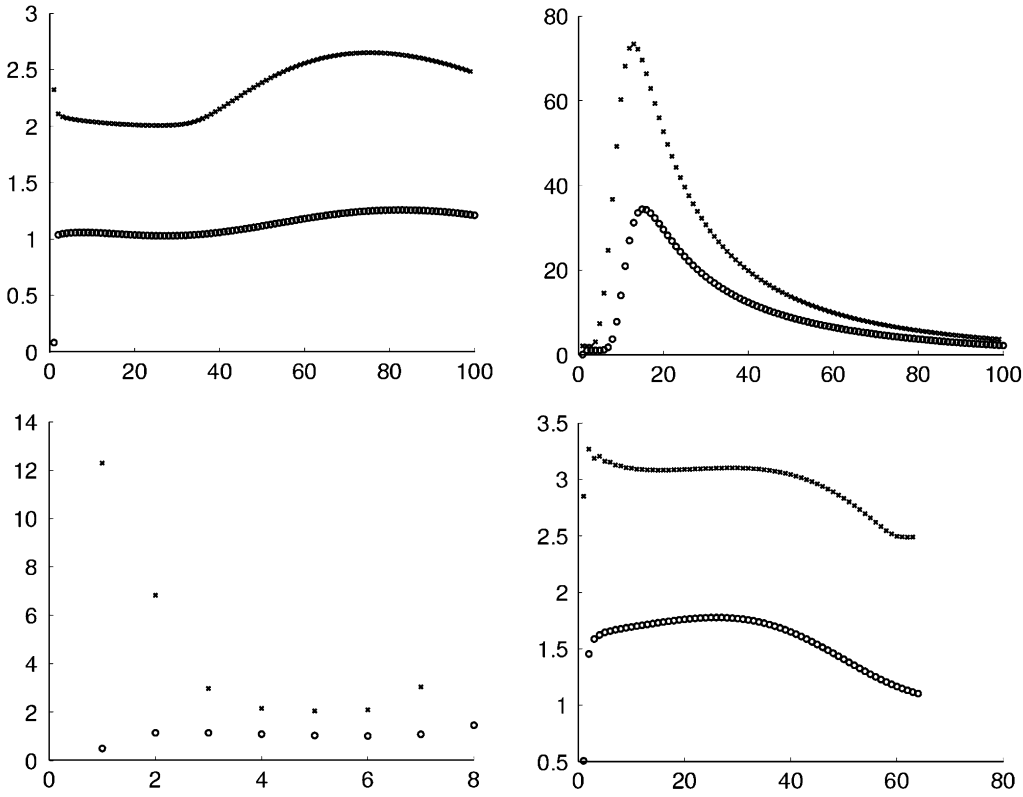

Fig. 2. Quotient $\lambda_{\max}(\tilde{L}_{rr}\tilde{L}_{rr}^{\mathrm{T}})/\lambda_{\min}(\tilde{L}_{n+1,n+1}\tilde{L}_{n+1,n+1}^{\mathrm{T}})$ (x) and norm $\left\|\tilde{L}_{rn}\tilde{L}_{nn}^{-1}\right\|$ (o) over n Example 1 (top left), Example 2 (top right), Example 3 (bottom left), Example 4 (bottom right).

convergence result for method 2(b) in Theorem 1 is confirmed, but also methods 1(a) and 2(a) exhibit convergence for these examples. Note that for the severely ill-posed Example 3 the singular values of T decay very fast, thus it does not make sense to consider more than the first eight columns of L or $\tilde{L}$, respectively.

To test this method also with noisy data and the discrepancy principle as a truncation rule, we add uniformly distributed random noise to the right-hand side. For each of the four examples and noise levels we made three experiments. The respective mean values of the stopping indices and the relative errors are listed in Table 1 and indicate convergence of the error as $\delta \to 0$. In all cases we used $\tau = 1.1$ in the discrepancy principle. It can be verified that in the second example $x^\dagger$ as given above satisfies the source condition (39) with $\nu = \frac{1}{2}$. As a matter of fact, the numbers in the second block of Table 1 indicate the convergence rate $O(\sqrt{\delta})$.

Finally, in Fig. 2 we show a numerical verification of the convergence conditions (38), (40) for method 2(b) with the discrepancy principle.

Remark 1. Concerning the numerical effort, the versions 1(b) and 2(b) have a twofold advantage: Only columns of length n (instead of "infinitely long" columns in the respective (a) versions) have to be computed. Moreover, the Cholesky factorization can be immediately used for computing the

application of $(L_{nn}L_{nn}^{\mathrm{T}})^{-1}$ (*or of* $(\tilde{L}_{nn}\tilde{L}_{nn}^{\mathrm{T}})^{-1}$) to some vector, by forward-backward substitution, as needed in the implementation. The latter is not the case for Methods 1(a) and 2(a), where $(L_nL_n^{\mathrm{T}})^{\dagger}$ (or $(\tilde{L}_n\tilde{L}_n^{\mathrm{T}})^{\dagger}$) has to be applied to some vector.

5. Conclusions and remarks

In this paper we have analyzed four tentative regularization methods for linear ill-posed operator equations, that are based on truncating the Cholesky factorization for positive definite matrices. We derived conditions for convergence in the noise free case and in case of noisy data, especially also with the discrepancy principle as truncation rule. These theoretical considerations and numerical test results as well as the computational effort clearly suggest one of the four methods as best, namely the one that is based on regularization by projection in data space and truncation of the factorization matrix up to its square upper left-hand part. Still we are left with some conditions that are hard to verify practically but have to be satisfied theoretically to guarantee convergence with noisy data. Realization of these conditions by means of appropriate matrix reordering strategies will be the subject of future research.

Acknowledgment

The author wishes to thank Prof. W. Hackbusch for interesting discussions that stimulated the idea of this paper.

Part of this work was completed during the authors stay at the Radon Institute for Computational and Applied Mathematics in Linz. Therefore, support by the Austrian Academy of Sciences is gratefully acknowledged.

Moreover, we wish to thank the referees for their valuable comments.

References

[1] M. Bebendorf, Hierarchical LU decomposition based preconditioners for BEM, Computing 74 (2005) 225–247.
[2] H.W. Engl, Integralgleichungen, Springer, Wien, 1997.
[3] H.W. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.
[4] B. Kaltenbacher, Regularization by projection with a posteriori discretization level choice for linear and nonlinear ill-posed problems, Inverse Problems 16 (2000) 1523–1539.
[5] B. Kaltenbacher, On the regularizing properties of truncated Cholesky factorization, RICAM-report No.06-9, Austrian Academy of Sciences, 2006.
[6] W. Hackbusch, A sparse matrix arithmetic based on matrices. Part I: introduction to $\mathscr{H}$-matrices, Computing 62 (1999) 89–108.
[7] S.V. Pereverzev, S. Prössdorf, On the characterization of self-regularization properties of a fully discrete projection method for Symm's integral equation, J. Integral Equations Appl. 12 (2) (2000) 113–130.
[8] R. Plato, Optimal algorithms for linear ill-posed problems yield regularization methods, Numer. Func. Anal. Optim. 11 (1990) 111–118.
[9] G. Vainikko, U. Hämarik, Projection methods and self-regularization in ill-posed problems, Sov. Math. 29 (1985) 1–20 (in Russian).